1990

Learning to Learn: The Art of Doing Science and Engineering, Session 28:

Hamming, Richard W.

Monterey, California: Naval Postgraduate School

Richard W. Hamming

Learning to Learn

The Art of Doing Science and Engineering

Session 28:
You Get What You Measure

Measurements & Organizations

The way you measure things has an effect on your organization & drawn conclusions

- Example: using nets to determine minimum size of fish in the sea

Example: Rating Systems

- Rating systems that rewards conservatism will remove risk-takers from the organization
- But risk-taking may be a trait that is needed later on

What You Choose to Measure

Hard to measure intelligence or morale

Confusion between what is reliably measured and what is relevant

- Tendency is to choose a thing that can be easily and accurately measured, versus hard-to-measure thing, without regard to relevance
- Adding reproducibility makes this choice harder still

Intelligence Quotient (IQ) Testing

Create a list of questions

- Test a small sample

Correlate question relevance to intelligence and drop “irrelevant” questions

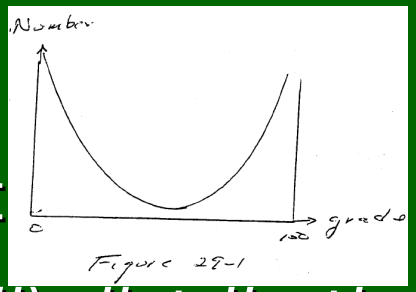
- Calibrate with a larger sample size

Forced IQs to be normally distributed through the calibration of the scores

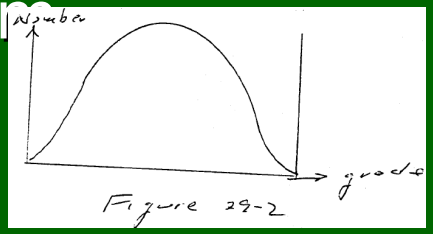
- irrespective of reality

Distribution of Grades

- **Final exam**
  - Questions can all be equally difficult
    - *Creates an all or nothing (pass/fail) distribution*



  - Some easy, some hard, most medium
    - *Creates a normal distribution*



- **Teacher can create whatever distribution desired**
  - Can even create test to fail a small group of students

Scoring Systems

Dynamic range (1-9 with 5 being the average)

- Most people will choose 4s and 6s
- One person can use 1s and 9s to dominate ratings
- Most people fail to use entire dynamic range

Scoring systems communicating information have maximum entropy when all symbols used equally

- Grading is a communication medium
- Giving all As and Bs provides little information
- Can adopt class rank to add info (but how good are peers?)

Rating People

- **Example: Bell Labs promotion and salary**
  - Rating people from different fields/departments
- **People do not like to rate people**
  - Judge not lest ye be judged; Cast not the first stone
- **Easier to determine relevant rank without giving the reason – the reason is where intuitive judgments are put into words**

Initially Perceived Features

The people you initially attract are the people you will later have

- Example: mixed up psychology students and faculty
- Example: CompSci – people obsessed with sea of detail

Causes inbreeding within field or company

- Strengthening most dominant perceived traits of organization/field (whether good or bad)
- Can weaken more subtle, “big picture” traits

Personnel Employment

- **Promote from within or go outside field**
- **Research needs people with original ideas**
  - These people may be “too original” for Human Resources (HR) recruiters
  - Company may need to get researchers to recruit other researchers (since like recognizes like)

Leadership & Promotions

- **Board of Directors self-selects leaders**
  - People they like and who were once like them, rather than people who will be good for the future
  - Great homogeneity leads to low innovation
  - High heterogeneity leads to no decisions being made
- **How to avoid inbreeding**
  - Don’t always choose someone from your own organization/field – once very common at universities
  - Think about how you are shaping the company and what would this all look like to an outsider

Judgements

- **Human vs. automated judgments**
  - "It's not that your answers are better than what we can do by hand, it is that they are consistent."
  - Systematic approach allowed study of subtle effects
  - Humans are better in taking the complexities of people and assigning them a scalar value (ranking)
- **Good human judgment requires maturity**
  - Example: to fail (or not fail) a failing student

Inspections

Random vs. scheduled

- People/organizations will prepare for inspections
- How does a scheduled evaluation relate to readiness at any given instant in time?
- While most “random” inspections are known in advance, it is usually not by as much as a scheduled inspection, thus providing a somewhat better opportunity to measure typical readiness

Scaling

More scales are available than just linear/additive.

Earthquakes measured on the logarithmic Richter scale (multiple of log of released energy).

- 2s & 3s common; 6s and 7s extremely rare
- Convenient to humans; Nature likely doesn't use logarithmic units to decide earthquake distribution

Logarithmic scale is good for many sensory tests.

Percentage change can be a good scale.

- Example: additional cattle into a herd (3 to 5 vs. 3 to 1000)

Decisions and Scaling

Scale is an important factor in making decisions and measuring/displaying data

- Equations will frequently do scaling

Lower mgt will bend figures for top mgt through creative scaling & measurement

- "How to Lie With Statistics" & "How to Lie with Charts"

- **Use due prudence to check figures/claims**
- Necessary for company health & your legal protection

Final Thoughts

Just because a measurement is popular, it does not make it reliable or accurate.

Capability does not equal probability.

- Underlings may bend those definitions
- Life testing measurements and tricks

Ask questions before creating a rating system

- What are the long term global effects?
- Who will we attract into our company?